

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : **10-301598**

**(43)Date of publication of application : 13.11.1998**

**G10L 3/02**

(71)Applicant : NIPPON HOSO KYOKAI <NHK>

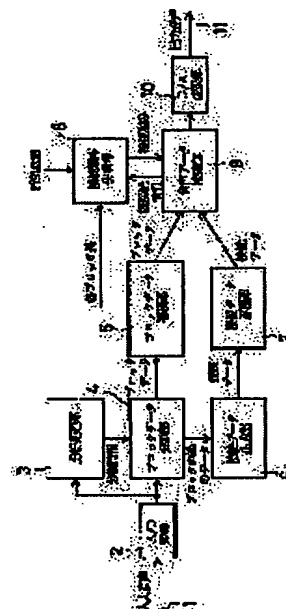
(72)Inventor : IMAI ATSUSHI  
SEIYAMA NOBUMASA  
TSUGI TORU

## (54) METHOD AND DEVICE FOR CONVERTING SPEECH SPEED

**(57)Abstract:**

**PROBLEM TO BE SOLVED:** To control adaptively conversion magnification of speech speed and a silence section according to set conditions and to stably obtain an effect expected in speech speed conversion in a really uttered time band by only once setting/operating one of several stages of the conversion magnification which is made as an aim for a user.

**SOLUTION:** When an uttering speed (speech speed) of a receiver voice is delayed, a connection order generation part 8 monitors always a data length of an input voice, an output data length beforehand calculated by a conversion function related to elongation/contraction magnification imparted beforehand and the data length of a really outputted voice at a fixed processing unit. Then, the connection order generation part 8 decides a connection order so as not to occur a contradiction in these, and controls a voice data connection part 9, and connects the voice data with the connection data without omitting voice information.



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-301598

(43) 公開日 平成10年(1998)11月13日

(51) Int.Cl.<sup>6</sup>

G 1 0 L 3/02

識別記号

F I

G 1 0 L 3/02

A

審査請求 有 請求項の数 8 O L (全 11 頁)

(21) 出願番号 特願平9-112961

(22) 出願日 平成9年(1997)4月30日

(71) 出願人 000004352

日本放送協会

東京都渋谷区神南2丁目2番1号

(72) 発明者 今井 篤

東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内

(72) 発明者 清山 信正

東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内

(72) 発明者 都木 徹

東京都世田谷区砧一丁目10番11号 日本放送協会放送技術研究所内

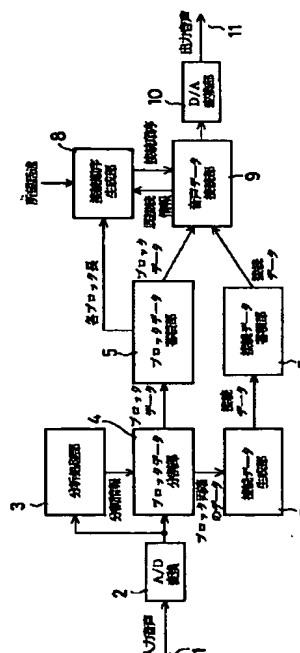
(74) 代理人 弁理士 三好 秀和 (外8名)

(54) 【発明の名称】 話速変換方法およびその装置

(57) 【要約】

【課題】 ユーザが数段階の目安となる変換倍率を一度だけ設定操作するだけで、設定された条件に応じて話速変換倍率や無音区間を適応的に制御し、実際に発話された時間枠の中で、話速変換に期待される効果を安定して得る。

【解決手段】 受聴音声の発声する速さ(話速)を遅くする際に、接続順序生成部8によって、入力音声のデータ長と、事前に与えられた伸縮倍率に関する変換関数によって予め計算された出力データ長と、実際に出力されている音声のデータ長とを一定の処理単位で常に監視し、これに矛盾が生じないように接続順序を決定し、音声データ接続部9を制御して、音声情報を欠落させることなく、音声データと、接続データとを接続させる。



## 【特許請求の範囲】

【請求項1】 時間的に変化する任意の比率で、入力データを伸張合成して得られた出力データについて、ある無音区間が出現し、この無音区間の継続時間が所定のしきい値を越えているとき、この入力データに対する出力データの伸張時間を、この伸張時間内の任意の時間だけ削減することを特徴とする話速変換方法。

【請求項2】 請求項1に記載の話速変換方法において、

入力データの伸縮合成する際、入力データ長と、この入力データ長に任意の伸縮倍率を乗じて算出される目標データ長と、実際の出力データ長との関係が矛盾しないように、逐次監視しながら、合成処理を行ない、時間的に変化する任意の伸縮合成比率に対し、音声部分に関して、情報の欠落が生じないようにするとともに、話速変換に伴う伸張に対する正確な時間情報を保持させることを特徴とする話速変換方法。

【請求項3】 請求項1または2に記載の話速変換方法において、

話速変換に伴う入力データ長からの伸張分を解消する際、一定継続時間以上の無音区間の一部を削除して、話速変換倍率、伸張量などに応じて、無音区間の残存割合を適応的に変化させることを特徴とする話速変換方法。

【請求項4】 請求項1、2、3のいずれかに記載の話速変換方法において、

限られた時間枠の中で、話速変換を行なう際、入力データ長と、この入力データ長に任意の伸縮倍率を乗じて算出される目標データ長と、実際の出力データ長との関係が矛盾しないように、逐次監視しながら、予め設定されている時間間隔で伸張量を測定し、この測定結果に基づき、時間差が少ないときには、話速変換倍率を一時的に上昇させ、また時間差が多いときには、話速変換倍率を一時的に下降させることにより、適応的に話速変換倍率を変化させることを特徴とする話速変換方法。

【請求項5】 入力データを各ブロックに分割してブロックデータを生成するとともに、各ブロックデータに基づき、接続データを生成する分割処理／接続データ生成手段と、

入力された所望話速に基づき、前記分割処理／接続データ生成手段によって生成された各ブロックデータ、各接続データの接続順序を決めて、これらを接続し、出力データを生成する接続処理手段とを備え、

この接続処理手段は、時間的に変化する任意の比率で、各ブロックデータを伸張合成して得られた出力データ中に無音区間が出現し、この無音区間の継続時間が所定のしきい値を越えているときには、このブロックデータに対する出力データの伸張時間を、この伸張時間内の任意の時間だけ削減することを特徴とする話速変換装置。

【請求項6】 請求項5に記載の話速変換装置において、

前記接続処理手段は、入力データを伸縮合成する際、入力データ長と、この入力データ長に任意の伸縮倍率を乗じて算出される目標データ長と、実際の出力データ長との関係が矛盾しないように、逐次監視しながら合成処理を行ない、

時間的に変化する任意の伸縮合成比率に対し、音声部分に関して情報の欠落が生じないようにするとともに、話速変換に伴う伸張に対する正確な時間情報を保持させることを特徴とする話速変換装置。

【請求項7】 請求項5または6に記載の話速変換装置において、

前記接続処理手段は、話速変換に伴う入力データ長からの伸張分を解消する際、一定継続時間以上の無音区間の一部を削除して、話速変換倍率、伸張量などに応じて、無音区間の残存割合を適応的に変化させることを特徴とする話速変換装置。

【請求項8】 請求項5、6、7のいずれかに記載の話速変換装置において、

前記接続処理手段は、限られた時間枠の中で、話速変換を行なう際、入力データ長と、この入力データ長に任意の伸縮倍率を乗じて算出される目標データ長と、実際の出力データ長との関係が矛盾しないように逐次監視しながら、予め設定されている時間間隔で伸張量を測定し、この測定結果に基づき、時間差が少ないときには、話速変換倍率を一時的に上昇させ、また時間差が多いときには、話速変換倍率を一時的に下降させることにより、適応的に話速変換倍率を変化させることを特徴とする話速変換装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、テレビジョン、ラジオ、テープレコーダ、ビデオテープレコーダ、ビデオディスクプレーヤ、補聴器などの映像機器、音響機器、医療機器などにおいて、時間を伸張させることなく、話速変換に期待される聞き易さを実現する話速変換方法およびその装置に関する。

【0002】【発明の概要】本発明は、人が発声した音声を加工してリアルタイムで発話速度を変換する話速変換方法およびその装置に関するものであって、受聴音声の発声する速さ（話速）を遅くする際に、入力音声のデータ長と、事前に与えられた伸縮倍率に関する変換関数によって予め計算された出力データ長と、実際に出力されている音声のデータ長とを一定の処理単位で常に監視しながら、情報の欠落を生じることなく、一連の処理を行なうものである。

【0003】さらに、この話速変換方法およびその装置においては、例えばテレビの視聴に使用する際、音声を伸張することによる映像と音声との時間差を最小限にすることを目的として、話速変換に期待される遅さの度合い（変換倍率）に応じて設定される可変のしきい値以上

の長さをも有する無音区間を適宜、短縮し、かつ入力データ長に対する出力データ長の時間差の程度によって適応的に変換倍率を変化させることにより、変換音声の発話時間を原音声の発話時間にほぼ保ちつつ、決められた時間枠の中で実現し得る最大のゆっくり感を自動的に生成するものである。

#### 【0004】

【従来の技術】話速変換手法を実際の放送に適用する場合、緊急報道など、原音声からの遅れが問題になる場合がある。特に、映像を伴うメディアについては、この遅れが話速変換に期待される効果とは逆に、悪影響を及ぼす可能性がある。

【0005】そこで、原音声から遅れることなく、話速変換効果（ゆっくり感）を実現する手法として、一様にゆっくり変換するのではなく、一息で行なう発声の開始点から終了点に向かう経過時間の関数として、話速をゆっくりから速くに変化させることで伸張を抑制し、文章間の無音区間を適宜、短縮する方法（池沢龍ほか、平成4年日本音響学会春期研究発表会「話速変換に伴う時間伸張を吸収するための一手法」2-6-2、pp. 331-332）や、この手法をリアルタイム処理化する方法（今井篤ほか、平成7年電子情報通信学会総合大会講演論文集「話速変換に伴う時間伸張のリアルタイム吸収法」D-694、pp. 300）などが報告されている。

【0006】前者は、全ての発話様式が既知として上で、適当な関数を手動で設定するものであり、後者も倍率を与える関数を手動で規定し、一度設定した後は、これを固定するものである。

【0007】一方、無音区間の短縮も、一定の残存時間のみを手動で規定するものであり、仮に「ずれ」が多く積算された場合には、バッファに蓄積された伸張分の音声を手動でクリアするものであった。

#### 【0008】

【発明が解決しようとする課題】このため、従来の話速変換装置では、放送音声の発話形態（話速や「間」のとり方など）が発話者によって様々であり、人手によって、それぞれに適したパラメータを設定しなければならないことから、操作箇所が多いとともに、設定自体が難しく、一般のユーザが取り扱うのに難し過ぎるという問題があった。

【0009】本発明は上記の事情に鑑み、請求項1～4では、ユーザが数段階の目安となる変換倍率を一度だけ設定操作するだけで、設定された条件に応じて話速変換倍率や無音区間を適応的に制御し、実際に発話された時間枠の中で、話速変換に期待される効果を安定して得ることができる話速変換方法を提供することを目的としている。

【0010】また、請求項5～8では、ユーザが数段階の目安となる変換倍率を一度だけ設定操作するだけで、

設定された条件に応じて話速変換倍率や無音区間を適応的に制御し、実際に発話された時間枠の中で、話速変換に期待される効果を安定して得ることができる話速変換装置を提供することを目的としている。

#### 【0011】

【課題を解決するための手段】上記の目的を達成するために、本発明による話速変換方法は、請求項1では、時間的に変化する任意の比率で、入力データを伸張合成して得られた出力データについて、ある無音区間が出現し、この無音区間の継続時間が所定のしきい値を越えているとき、この入力データに対する出力データの伸張時間を、この伸張時間内の任意の時間だけ削減することを特徴としている。

【0012】また、請求項2では、請求項1に記載の話速変換方法において、入力データの伸縮合成する際、入力データ長と、この入力データ長に任意の伸縮倍率を乗じて算出される目標データ長と、実際の出力データ長との関係が矛盾しないように、逐次監視しながら、合成処理を行ない、時間的に変化する任意の伸縮合成比率に対し、音声部分に関して、情報の欠落が生じないようにするとともに、話速変換に伴う伸張に対する正確な時間情報を保持させることを特徴としている。

【0013】また、請求項3では、請求項1または2に記載の話速変換方法において、話速変換に伴う入力データ長からの伸張分を解消する際、一定継続時間以上の無音区間の一部を削除して、話速変換倍率、伸張量などに応じて、無音区間の残存割合を適応的に変化させることを特徴としている。

【0014】また、請求項4では、請求項1、2、3のいずれかに記載の話速変換方法において、限られた時間枠の中で、話速変換を行なう際、入力データ長と、この入力データ長に任意の伸縮倍率を乗じて算出される目標データ長と、実際の出力データ長との関係が矛盾しないように、逐次監視しながら、予め設定されている時間間隔で伸張量を測定し、この測定結果に基づき、時間差が少ないときには、話速変換倍率を一時的に上昇させ、また時間差が多いときには、話速変換倍率を一時的に下降させることにより、適応的に話速変換倍率を変化させることを特徴としている。

【0015】また、上記の目的を達成するために、本発明による話速変換装置は、請求項5では、入力データを各ブロックに分割してブロックデータを生成するとともに、各ブロックデータに基づき、接続データを生成する分割処理／接続データ生成手段と、入力された所望話速に基づき、前記分割処理／接続データ生成手段によって生成された各ブロックデータ、各接続データの接続順序を決めて、これらを接続し、出力データを生成する接続処理手段とを備え、前記接続処理手段は、時間的に変化する任意の比率で、各ブロックデータを伸張合成して得られた出力データ中に無音区間が出現し、この無音区間

の継続時間が所定のしきい値を越えているとき、このブロックデータに対する出力データの伸張時間を、この伸張時間内の任意の時間だけ削減することを特徴としている。

【0016】また、請求項6では、請求項5に記載の話速変換装置において、前記接続処理手段は、入力データの伸縮合成する際、入力データ長と、この入力データ長に任意の伸縮倍率を乗じて算出される目標データ長と、実際の出力データ長との関係が矛盾しないように逐次監視しながら、合成処理を行ない、時間的に変化する任意の伸縮合成比率に対し、音声部分に関して、情報の欠落が生じないようにするとともに、話速変換に伴う伸張に対する正確な時間情報を保持させることを特徴としている。

【0017】また、請求項7では、請求項5または6に記載の話速変換装置において、前記接続処理手段は、話速変換に伴う入力データ長からの伸張分を解消する際、一定継続時間以上の無音区間の一部を削除して、話速変換倍率、伸張量などに応じて、無音区間の残存割合を適応的に変化させることを特徴としている。

【0018】また、請求項8では、請求項5、6、7のいずれかに記載の話速変換装置において、前記接続処理手段は、限られた時間枠の中で、話速変換を行なう際、入力データ長と、この入力データ長に任意の伸縮倍率を乗じて算出される目標データ長と、実際の出力データ長との関係が矛盾しないように、逐次監視しながら、予め設定されている時間間隔で伸張量を測定し、この測定結果に基づき、時間差が少ないときには、話速変換倍率を一時的に上昇させ、また時間差が多いときには、話速変換倍率を一時的に下降させることにより、適応的に話速変換倍率を変化させることを特徴としている。

【0019】上記の構成において、請求項1に記載の話速変換方法では、時間的に変化する任意の比率で、入力データを伸張合成して得られた出力データについて、ある無音区間が出現し、この無音区間の継続時間が所定のしきい値を越えているとき、この入力データに対する出力データの伸張時間を、この伸張時間内の任意の時間だけ削減することにより、ユーザが数段階の目安となる変換倍率を一度だけ設定操作するだけで、設定された条件に応じて話速変換倍率や無音区間を適応的に制御し、実際に発話された時間枠の中で、話速変換に期待される効果を安定して得る。

【0020】また、請求項2では、入力データを伸縮合成する際、入力データ長と、この入力データ長に任意の伸縮倍率を乗じて算出される目標データ長と、実際の出力データ長との関係が矛盾しないように、逐次監視しながら、合成処理を行ない、時間的に変化する任意の伸縮合成比率に対し、音声部分に関して、情報の欠落が生じないようにするとともに、話速変換に伴う伸張に対する正確な時間情報を保持させることにより、ユーザが数段

階の目安となる変換倍率を一度だけ設定操作するだけで、設定された条件に応じて話速変換倍率や無音区間を適応的に制御し、実際に発話された時間枠の中で、話速変換に期待される効果を安定して得る。

【0021】また、請求項3では、話速変換に伴う入力データ長からの伸張分を解消する際、一定継続時間以上の無音区間の一部を削除して、話速変換倍率、伸張量などに応じて、無音区間の残存割合を適応的に変化させることにより、ユーザが数段階の目安となる変換倍率を一度だけ設定操作するだけで、設定された条件に応じて話速変換倍率や無音区間を適応的に制御し、実際に発話された時間枠の中で、話速変換に期待される効果を安定して得る。

【0022】また、請求項4では、限られた時間枠の中で、話速変換を行なう際、入力データ長と、この入力データ長に任意の伸縮倍率を乗じて算出される目標データ長と、実際の出力データ長との関係が矛盾しないように、逐次監視しながら、予め設定されている時間間隔で伸張量を測定し、この測定結果に基づき、時間差が少ないときには、話速変換倍率を一時的に上昇させ、また時間差が多いときには、話速変換倍率を一時的に下降させることにより、適応的に話速変換倍率を変化させることにより、ユーザが数段階の目安となる変換倍率を一度だけ設定操作するだけで、設定された条件に応じて話速変換倍率や無音区間を適応的に制御し、実際に発話された時間枠の中で、話速変換に期待される効果を安定して得る。

【0023】また、請求項5に記載の話速変換装置では、入力データを各ブロックに分割してブロックデータを生成するとともに、各ブロックデータに基づき、接続データを生成する分割処理／接続データ生成手段と、入力された所望話速に基づき、前記分割処理／接続データ生成手段によって生成された各ブロックデータ、各接続データの接続順序を決めて、これらを接続し、出力データを生成する接続処理手段とを有する話速変換装置において、前記接続処理手段によって、時間的に変化する任意の比率で、各ブロックデータを伸張合成して得られた出力データ中に無音区間が出現し、この無音区間の継続時間が所定のしきい値を越えているとき、このブロックデータに対する出力データの伸張時間を、この伸張時間内の任意の時間だけ削減することにより、ユーザが数段階の目安となる変換倍率を一度だけ設定操作するだけで、設定された条件に応じて話速変換倍率や無音区間を適応的に制御し、実際に発話された時間枠の中で、話速変換に期待される効果を安定して得る。

【0024】また、請求項6では、前記接続処理手段によって、入力データの伸縮合成する際、入力データ長と、この入力データ長に任意の伸縮倍率を乗じて算出される目標データ長と、実際の出力データ長との関係が矛盾しないように、逐次監視しながら、合成処理を行な

い、時間的に変化する任意の伸縮合成比率に対し、音声部分に関して、情報の欠落が生じないようにするとともに、話速変換に伴う伸張に対する正確な時間情報を保持させることにより、ユーザが数段階の目安となる変換倍率を一度だけ設定操作するだけで、設定された条件に応じて話速変換倍率や無音区間を適応的に制御し、実際に発話された時間枠の中で、話速変換に期待される効果を安定して得る。

【0025】また、請求項7では、前記接続処理手段によって、話速変換に伴う入力データ長からの伸張分を解消する際、一定継続時間以上の無音区間の一部を削除して、話速変換倍率、伸張量などに応じて、無音区間の残存割合を適応的に変化させることにより、ユーザが数段階の目安となる変換倍率を一度だけ設定操作するだけで、設定された条件に応じて話速変換倍率や無音区間を適応的に制御し、実際に発話された時間枠の中で、話速変換に期待される効果を安定して得る。

【0026】また、請求項8では、前記接続処理手段によって、限られた時間枠の中で、話速変換を行なう際、入力データ長と、この入力データ長に任意の伸縮倍率を乗じて算出される目標データ長と、実際の出力データ長との関係が矛盾しないように、逐次監視しながら、予め設定されている時間間隔で伸張量を測定し、この測定結果に基づき、時間差が少ないときには、話速変換倍率を一時的に上昇させ、また時間差が多いときには、話速変換倍率を一時的に下降させることにより、適応的に話速変換倍率を変化させることにより、ユーザが数段階の目安となる変換倍率を一度だけ設定操作するだけで、設定された条件に応じて話速変換倍率や無音区間を適応的に制御し、実際に発話された時間枠の中で、話速変換に期待される効果を安定して得る。

【0027】

【発明の実施の形態】図1は本発明の実施の形態としての話速変換装置の一例を示すブロック図である。

【0028】この図に示す話速変換装置は、端子1と、A/D変換部2と、分析処理部3と、ブロックデータ分割部4と、ブロックデータ蓄積部5と、接続データ生成部6と、接続データ蓄積部7と、接続順序生成部8と、音声データ接続部9と、D/A変換部10と、端子11とを備えており、発話者からの入力音声データに対し、音声データの属性に基づく分析処理を施し、当該分析情報に応じて所望の関数を使用して、話速変換音声データを合成する際、入力音声データのデータ長（入力データ長）と、これに任意の伸縮倍率を乗じて算出される＊

$$P_{thr} = P_{upper} - 35$$

$P_{upper} - P_{lower} < 60$  [dB] の場合、

$$P_{thr} = P_{upper} - 35 + 35 \{1 - (P_{upper} - P_{lower}) / 60\} \dots (2)$$

但し、 $P_{thr} : P_{thr} = P_{upper} - 13$  を上限とする。

【0035】そして、パワーが所定のしきい値  $P_{thr}$  以上の区間については、声帯の振動を伴う音声である有

＊目標データ長と、実際の出力音声データのデータ長（出力データ長）とを比較しながら、矛盾がないように、これらの処理を行なうことにより、伸張・伸縮倍率の変化に対しても、音声情報の欠落が生じることが無く、また時々刻々、変化する原音声と、変換音声との時間差を監視する。そして、時間差が少ない場合には、話速変換倍率を一時的に上昇させ、また逆に多い場合には、話速変換倍率を一時的に下降させなど、適応的に倍率を変化させ、さらに話速変換倍率や伸張量などに基づいて、無音区間の残存割合を適応的に変化させて、話速変換に伴う原音声からの時間差を適応的に解消する。

【0029】A/D変換部2では、所定のサンプリングレート（例えば、32kHz）で、端子1に入力された音声信号、例えばマイクロホンやテレビジョン、ラジオ、その他の映像機器、音響機器のアナログ音声出力端子から出力される音声信号をA/D変換するとともに、これによって得られた音声データをFIFOメモリにバッファリングしながら、過不足なく、後続の分析処理部3と、ブロックデータ分析部4とに供給する。

【0030】分析処理部3では、A/D変換部2から出力される音声データを分析して、各属性を抽出するとともに、これらの各属性に基づいて、ブロックデータ分割部4において行われる音声データの分割処理に必要な各ブロック時間長を決定する分割情報を生成し、これをブロックデータ分割部4に供給する。

【0031】この際、入力された音声データの各属性として、有声音、無声音、無音を設定する。雑音や音楽などの背景音といった属性も考えられるが、一般に、雑音や背景音の信号と、音声信号とを自動的に判別することが難しいことから、雑音、背景音も、上述した3つの属性の1つに分類して分析を行なう。

【0032】まず、分析処理部3における処理量を低減するために、サンプリングレートを16kHzまで落とすデシメーションを施す。次に、5ms前後の間隔で、音声データのパワーを算出するため、10ms前後の窓幅において、データの自乗和を計算する。このパワーが所定のしきい値  $P_{thr}$  未満の場合、その部分を無音区間に決定する。

【0033】パワーしきい値の決定は、瞬時パワーの最大値  $P_{upper}$  および瞬時パワーの最小値  $P_{lower}$  を用いて、次式に示すように、パワーに関するしきい値  $P_{thr}$  を決定する。

【0034】

$$\text{【数1】 } P_{upper} - P_{lower} \geq 60 \text{ [dB] の場合、} \dots (1)$$

音か、声帯の振動を伴わない音声である無声音かの判定を行なう。これには、パワーの大きさだけでなく、ゼロ交差分析、自己相関分析などを併用する。

【0036】また、音声データを分析するために、各ブロックの時間長を決定するときには、各属性毎に所定の自己相関分析を行なって周期性を検出し、この周期性を基に、ブロック長を決定する。また、有声音区間については、声帯の振動周期であるピッチ周期を検出し、各ピッチ周期が各々のブロック長となるように分割を行なう。この際、有声音区間のピッチ周期が1.25ms～28.0ms程度の広い範囲に分布しているため、長短異なる窓幅の自己相関分析を行なうなどして、できるだけ正確なピッチ周期を検出する。なお、有声音区間のブロック長として、ピッチ周期を用いたのは、ブロック単位の繰り返しに起因する声の高さの変化（低い声になる）を防止するためである。また、無声音区間、無音区間については、5ms以内の周期性を検出して、ブロック長を検出する。

【0037】また、ブロックデータ分割部4では、分析処理部3で決定されたブロック長にしたがって、A/D変換部2から出力される音声データを分割し、この分割処理で得られたブロック単位の音声データと、そのブロック長とをブロックデータ蓄積部5に供給するとともに、分割処理で得られた各ブロック単位の音声データの両端部分、すなわち開始部分から所定の時間長（例えば、2ms分）と、終了部分から所定の時間長（例えば、2ms分）前の部分を接続データ生成部6に供給する。

【0038】また、ブロック蓄積部5では、リングバッファによって、ブロックデータ分割部4から供給されたブロック単位の音声データ、そのブロック長を一時的に格納し、必要に応じて一時記憶しているブロック単位の音声データを音声データ接続部6に供給するとともに、必要に応じて一時記憶しているブロック長を接続順序生成部8に供給する。

【0039】また、接続データ生成部6では、各ブロック毎に、図2に示すように、直前のブロックの終了部分、当該ブロックの開始部分の音声、直後のブロックの開始部分の音声データに窓掛けを行なった後、直前のブロックの終了部分と、当該ブロックの終了部分の重複加算および当該ブロックの開始部分と直後のブロック開始部分の重複加算を行なうとともに、これらを連結して各ブロック毎に、接続データを生成し、これを接続データ蓄積部7に供給する。

【0040】接続データ蓄積部7では、リングバッファによって、接続データ生成部6から供給された各ブロック毎の接続データを一時記憶するとともに、必要に応じて一時記憶している接続データを音声データ接続部9に供給する。

【0041】また、接続順序生成部8では、受聴者が設定した所望の話速を実現するために、ブロック単位の音声データおよび接続データの接続順序を生成する。この場合、受聴者がデジタルボリュームなどをインタフェー

スとして、各属性毎の時間的な伸張倍率を設定できる。この値は書き換え可能なメモリに格納されている。またこの値は、固定の伸張倍率として処理される方法（＝一様伸張モード）と、この設定倍率を目標にしつつ、一定時間以上ずれが積算しないように、各音声属性を総合的に、かつ適応的に制御することで、限られた時間枠で話速変換効果を実現する方法（＝時間伸張吸収モード）とのいずれかを選択することによって提供される。

【0042】この接続順序生成部8によれば、上記メモリに設定された伸張倍率に対して実際に音声合成を行なう際に、同時刻の入力音声データ長と出力音声データ長と、これから合成しようとする音声データ長の各時間関係をリアルタイムで把握することで、原音声の発話時刻と変換音声の出力時刻との時間差を常に監視することができ、この情報をフィードバックすることで時間差を自動的に一定長以下に抑え込むことができる。また同時に、任意のタイミングで任意の値に変更される伸縮倍率に対して、その実行に時間的な矛盾（例えば、入力音声データ長よりも出力音声データ長を短くするような要求など）がないか否かをチェックでき、合成時に音声情報の欠落を生ずることを防止できる。

【0043】次に、この接続順序生成部8の処理を具体的に説明する。任意の関数によって音声の伸縮倍率を設定する際、ブロックデータ蓄積部5から供給される各ブロック長に基づき、ブロックデータ分割部4で規定された処理単位の音声データ長（＝入力データ長）を逐次算出し、この入力データ長に対し、受聴者によって設定された伸縮倍率を乗じたものを目標データ長とする。音声データ接続部9では、この目標データ値と一致するように音声データを接続するとともに、実際に出力された出力音声データの長さとなる音声データ長（＝出力データ長）を逐次、接続順序生成部8にフィードバックする。

【0044】そして、図3に示すように、接続順序生成部8に設けられた入出力データ長監視比較部20、即ち入力データ長を監視する入力データ長監視部21と、この入力データ長監視部21で得られた入力データ長と例えば受聴者（あるいは、装置に内蔵された関数メモリ）によって与えられた値に基づいて行われた話速倍率変換で生成される出力データの目標長（目標データ長）を演算するとともに、この目標データ長を自動的に修正する出力目標長演算部22と、この出力目標長演算部22で得られた目標データ長と入力データ長監視部21で得られた入力データ長とを比較して、目標データ長が入力データ長より短いときは目標データ長を入力データ長に揃え、さらに、目標データ長が入力データ長以上のときは目標データ長をそのまま出力する比較部23と、音声データ接続部9から出力データに関する既接続情報を入力して出力データ長を監視する出力データ長監視部24と、この出力データ長監視部24で得られた出力データ長と比較部23で得られた目標データ長とを比較し、目

標データ長が出力データ長より短いときは目標データ長を出力データ長に揃え、さらに、目標データ長が出力データ長以上のときは目標データ長をそのまま出力する比較部25とによって生成される目標長を、接続順序情報として音声データ接続部9に送る。そして、次に述べるように、音声の属性毎に設定されたメモリの値を所定の時間間隔で読み出すとともに、読み出された属性毎の伸張倍率を実現するために、目標データ長を求めるとも、この目標データ長と、出力データ長監視部24で得られた出力データ長に基づき、音声の伸縮情報を加味した接続情報を時々刻々、生成して、図4に示すように、各ブロック毎の音声データと、接続データとを接続させる。

【0045】まず、入力データ長と、目標データ長とを逐次比較し、入力データ長が目標データ長以上と判定されたときには、入力データ長に揃うように、目標データ長を修正し、また入力データ長が目標データ長未満であると判定されたときには、目標データ長の変更を中止する。

【0046】次に、目標データ長と、実際の出力データ長とを比較し、出力データ長が目標データ長以上と判定されたときには、出力データ長に揃うように、目標データ長を修正し、また出力データ長が目標データ長未満と\*

$$f(t) = rs + 0.5(rs - re)(\cos \pi t / T + 1, 0) \dots (3)$$

但し、 $t: 0 \leq t \leq T$

$rs$ : 受聴者による外部入力値 ( $1.0 \leq rs \leq 1.6$ )

$re$ : 初期値として与えられる値 (例えば、 $re = 1.0$ )

【0051】ここで、入力データ長と、出力データ長との時間差をある一定の時間間隔、例えば1秒毎に計算し、そのときの時間差に応じて、初期値 $re$ を“1.0”から“0.05”ずつ増加させたり、また逆に“0.95”程度まで減少させる処理を行なう。ただし、期間 $T$ を越えた時点で、まだ200ms以上の無音区間が出現しない場合には、それ以降の有声区間には、例えば1.0倍の倍率を適用する。ここでは、ピッチやパワーなどの変化量を指標にして新たな倍率を与えることもできる。

【0052】また、無音区間の残存割合についても、話速変換倍率や伸張量などを鑑みて適応的に変化させるようにする。これも関数として任意に設定できる。

【0053】また、外部入力値 $rs$ に対応して無音区間の短縮許容限 (最低、どれだけは削減せずに保存するかを示す値)を設定し、上述したような関数で表現しても良いが、例えば次に述べるように、離散的に設定することもできる。

【0054】

$rs = 1.0$ のときは、300msまで削減可能

$rs = 1.1$ のときは、250msまで削減可能

\* 判定されたときには、目標データ長の変更を中止する。

【0047】これらの比較処理によって得られた目標データ長と合致するように、伸張情報や接続情報などを示す接続指令を生成して、これを音声データ接続部9に供給する。

【0048】次に、接続順序生成部8における話速変換倍率の制御条件について説明する。例えば、放送の時間枠など、限られた時間枠の中で、話速変換を行なうことを所望する場合においては、入力データ長と、出力データ長とを逐次監視し、予め任意に設定した時間間隔で、両データの時間差を測定することによって、遅延量が少ないときには、話速変換倍率を一時的に上昇させ、また逆に多いときには、これを下降させる処理を行なうなど、適応的に倍率を変化させるような関数を設定すれば良い。

【0049】例えば、この実施の形態では、200ms以上の無音区間が出現した時点で、それ以降に出現する最初の有声音の開始時刻を“ $t = 0$ ”とし、“ $0 \leq t \leq T$ ”の範囲に出現する各有声音の開始時刻に対応した倍率を与える関数として、以下のような余弦関数を用いることができる。

【0050】

【数2】

$$f(t) = rs + 0.5(rs - re)(\cos \pi t / T + 1, 0) \dots (3)$$

$rs = 1.2$ のときは、230msまで削減可能

$rs = 1.3$ のときは、200msまで削減可能

$rs = 1.4$ のときは、200msまで削減可能

$rs = 1.5$ のときは、150msまで削減可能

$rs = 1.6$ のときは、100msまで削減可能

などに設定しても良い。

【0055】また、無音区間の削減方式については、リングバッファ上の任意のアドレスにポインタを移動させることによって実現する。この実施の形態では、当該無音区間の直後の有声音の開始部分に移動することにより、音声情報の欠落を防止している。

【0056】また、音声データ接続部9では、接続順序生成部8で決定された接続順序にしたがってブロックデータ蓄積部5からブロック単位の音声データを読み出し、指定されたブロックの音声データを伸張させるとともに、接続データ蓄積部7から接続データを読出しながら、D/A変換部10に設けられたFIFOメモリに過不足が起らないように、接続処理を抑制しながら、音声データと接続データとを接続して、出力音声データを生成し、これをD/A変換部10に供給する。

【0057】D/A変換部10では、FIFOメモリによって、音声データ接続部9から供給される出力音声データをバッファリングしながら、所定のサンプリングレート (例えば、32kHz) で、出力音声データをD/A変換して、出力音声信号を生成し、これを端子11から出力する。



【0058】このように、この実施の形態では、発話者からの入力音声データに対して、音声データの属性に基づく分析処理を施し、当該分析情報に応じた所望の関数を使用して話速変換音声データを合成する際、入力データ長と、これに任意の伸縮倍率を乗じて算出される目標データ長と、実際の出力音声データ長とを比較しながら、矛盾がないように、これらの処理を行なうようにしたので、伸張・伸縮倍率の変化に対しても、音声情報の欠落が生じないようにすることができる。また、時々刻々変化する原音声と、変換音声との時間差を監視し、時間差が少ない場合には、話速変換倍率を一時的に上昇させ、また逆に多い場合には、話速変換倍率を一時的に下降させるなど、適応的に倍率を変化させ、さらに話速変換倍率や伸張量などに基づいて、無音区間の残存割合を適応的に変化させて、話速変換に伴う原音声からの時間差を適応的に解消するようにしているので、ユーザが数段階の目安となる変換倍率を一度だけ設定操作するだけで、設定された条件に応じて話速変換倍率や無音区間を適応的に制御し、実際に発話された時間枠の中で、話速変換に期待される効果を安定して得ることができる。

【0059】これによって、話者が頻繁に入れ替わる放送番組などに対しても、自動的に各発話者に最適な話速変換効果を提供することができ、ごく簡単な操作で、早口が聞き取り難い高齢者や視聴障害者に対しても、リアルタイム性を有する緊急報道やテレビなどの映像付きのメディアの音声を時間遅れ無く、かつ安定してゆっくり聴取させることができる。

#### 【0060】

【発明の効果】以上説明したように本発明によれば、請求項1～4では、ユーザが数段階の目安となる変換倍率を一度だけ設定操作するだけで、設定された条件に応じて話速変換倍率や無音区間を適応的に制御し、実際に発話された時間枠の中で、話速変換に期待される効果を安定して得ることができる。

【0061】また、請求項5～8では、ユーザが数段階の目安となる変換倍率を一度だけ設定操作するだけで、設定された条件に応じて話速変換倍率や無音区間を適応

的に制御し、実際に発話された時間枠の中で、話速変換に期待される効果を安定して得ることができる。

#### 【図面の簡単な説明】

【図1】本発明を適用した話速変換装置の一例を示すブロック図である。

【図2】図1に示す接続データ生成部における、同一ブロックを繰り返して接続する際に用いる接続データの生成方法を示す模式図である。

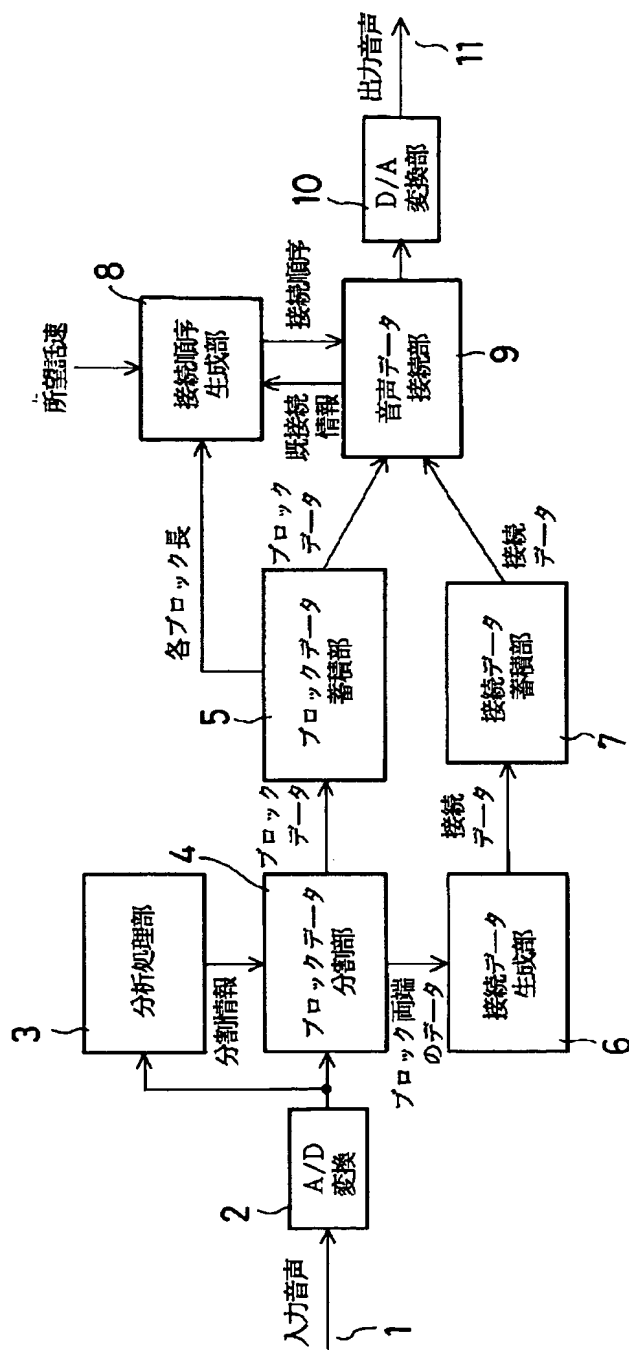
【図3】図1に示す接続順序生成部における入出力データ長監視比較部の詳細な構成例を示すブロック図である。

【図4】図1に示す接続順序生成部で生成される接続順序の一例を示す模式図である。

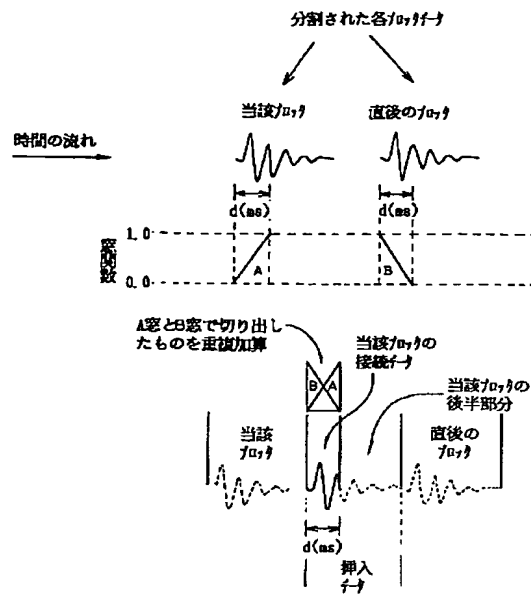
#### 【符号の説明】

- 1 端子
- 2 A/D変換部
- 3 分析処理部（分割処理／接続データ生成手段）
- 4 ブロックデータ分割部（分割処理／接続データ生成手段）
- 5 ブロックデータ蓄積部（分割処理／接続データ生成手段）
- 6 接続データ生成部（分割処理／接続データ生成手段）
- 7 接続データ蓄積部（分割処理／接続データ生成手段）
- 8 接続順序生成部（接続処理手段）
- 9 音声データ接続部（接続処理手段）
- 10 D/A変換部
- 11 端子
- 12 話速変換装置
- 20 入出力データ長監視比較部
- 21 入力データ長監視部
- 22 出力目標長演算部
- 23 比較部
- 24 出力データ長監視部
- 25 比較部

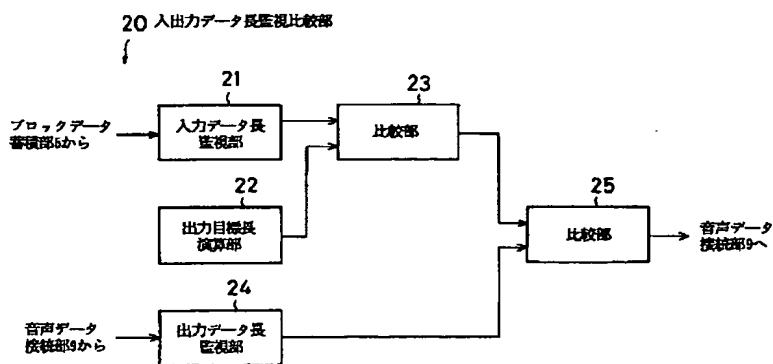
【図1】



【図2】



【図3】



【図4】

